

## **"Evaluation of prediction models: from AUC to calibration and decision curve analysis" (half day)**

Ewout Steyerberg & Ed Bonnevill, LUMC Leiden; Ben Van Calster, KU Leuven

Prediction models relate multiple patient or disease characteristics to diagnostic or prognostic outcomes. Prediction models enjoy increasing popularity with our increasing knowledge on markers, imaging and hopes for Artificial Intelligence / Machine Learning as flexible modeling tools. Validation of predictions from such models is essential, whether developed by classical statistical methods or novel AI approaches. Evaluation of performance commonly starts with an assessment of the discriminative ability of a prediction model, as often quantified by the Area under ROC curve (AUC) for binary outcomes. Other performance measures are popular among the AI community, such as the F1 score and Precision Recall Curve (PRC). In addition to discrimination, calibration is key to assess when predictions intend to be communicated to clinicians and patients, and support shared decision-making. Various summary measures are available to the analyst to complement graphical calibration assessment. Nowadays, further evaluation includes the Net Benefit, a summary measure that depends on the risk threshold for defining high versus low risk. Net Benefit is related to the threshold in a Decision Curve Analysis (DCA).

### **Course outline**

1. Overview of traditional and modern performance measures for binary outcomes
2. Detailed overview of decision curve analysis to assess potential clinical utility
3. Discussion of different types of model validation, including choice of performance metrics
4. Extension of performance measures to nominal, ordinal, survival, and competing risk outcomes

### **Learning goals**

- Obtain an up-to-date understanding of statistical performance measures for prediction models, and extensions to decision-analytic measures
- Know about measures for different types of outcomes, beyond binary outcomes: categorical; ordinal outcomes; survival; competing risks
- Learn how to estimate apparent, internally validated, and externally validated performance
- Be able to apply approaches in R

### **Target audience**

Applied biostatisticians interested in prediction research, diagnostic and prognostic modeling, validation concepts